Mobile Edge Computing for 5G: The Communication Perspective

Kaibin Huang
Dept. of Electrical & Electronic Engineering
The University of Hong Kong
Hong Kong

Joint Work with Yuyi Mao (HKUST), Changsheng You (HKU), Jun Zhang (HKUST), Khaled B. Letaief (HKUST)
be able to provide intelligent optimization based on services and users awareness, and will improve energy and cost efficiency by over a hundred of times, enabling us all to realize the vision of 5G - "Information a finger away, everything in touch".

5G Overall Vision

Communication, **Computing** and Content Delivery

- Tactile-scale latency (milliseconds)
- Ubiquitous
- Gigabit wireless access
- Secure
- Device-to-device
- High mobility
Mobile Edge Computing (MEC)

• MEC provides cloud-computing capabilities at the edge of the mobile network (base stations, access points, set-top boxes…) in close proximity to mobile subscribers.

• The environment of MEC is characterized by low latency, proximity, high bandwidth, real-time, location and context awareness.
MEC Application: Face Recognition

- Offload **computation-intensive** components to MEC server

![Diagram showing the process of face recognition](image)

- **Image acquisition**
  - Digital image
  - Classification results

- **Face detection**
  - Face image

- **Pre-processing**
  - Normalized Face image

- **Feature extraction**
  - Feature vector

- **Classification**

**Database** (training samples)

**MEC server**

Should be computed locally
MEC Application: Augmented Reality

- Offload **computation-intensive** components to MEC server

Should be computed locally
Increasing propagation latency vs. reducing computation latency
Mobile Computation Offloading (MECO)

- Mobile computation offloading can save energy when the following inequality is satisfied

\[ p_m \frac{w}{s_m} > p_c \frac{d_i}{B} + p_i \frac{w}{s_s} \]

- \( w \): workload
- \( s_m \): computation speed of the mobile
- \( d_i \): data size
- \( B \): bandwidth
- \( p_m \): computation power of the mobile
- \( p_c \): transmit power of the mobile
- \( p_i \): idle state power of the mobile
- \( s_s \): computation speed of the cloud
Integrating Computer Science and Wireless Communication

Wireless Communication
- Radio resource management
- Multiple access control
- Physical layer techniques (Massive MIMO, mmWave…)
- Small-cell networks
- …

Computer Science
- Network function virtualization
- Computing offloading
- Cloud computing
- Software architecture
- Parallel computing
- Live migration
- …
Resource Management in MEC Systems
Joint Radio-and-Computational Resource Allocation

Radio resources

Computational resources

- Centralized resource allocation
- Distributed resource allocation

Channel? What to compute?

Save energy
Centralized framework: TDMA system, limited-capacity MEC server

- **Offloading priority function** on local computing energy consumption, channel gain and data size
- Optimal policy structure: **Threshold-based** binary offloading decision

Policy Structure

- **Descending order**
- **Offloading Priority**
- **Resource allocation**
  - Offloading time-sharing duration
  - Cloud computation capacity

---

MEC Server Scheduling

Asynchronous arrival

Varying Latency requirements
Sensitive → Tolerant

Task Dependency for Individual Users

Image recognition
Source: http://ochoa.cs.cmu.edu/wschu/project-fr.html

Tools: queuing theory, integer programming
Multiuser Cooperative Edge Computing

Computational resources

Detection

D2D

Overload!

Balance
Peer-to-Peer Mobile Cooperative Computing

- Motivation: Scavenging nearby **real-time idling** computation resources

![Diagram](https://via.placeholder.com/500)

**An energy efficiency dilemma:** Longer time for transmission reduce TX energy but scavenge less peer CPU resources, and vice versa

---

Issues, Challenges, and Research Opportunities
will be asynchronously and repeatedly requested. Motivated by Cisco’s prediction, mobile video streaming will occupy up to 72% of the entire mobile data traffic by 2019. However, different mobile services require different resources, e.g., the Ginibre stochastic geometry theory. Such analysis of MEC systems should address the following challenges: 1) The timescales of computation and wireless channel coherence time may be different [81], [104], which makes existing results for conventional Cloud Computing systems [162]–[165], termed as MCC systems [161]. Specifically, the timescales of mobile devices requesting for MEC services will be more concentrated around the MEC servers. It has been predicted by Cisco that mobile video streaming requests are highly concentrated and some popular contents will be asynchronously and repeatedly requested. Motivated by Cisco’s prediction, mobile video streaming will occupy up to 72% of the entire mobile data traffic by 2019.

One possible solution is to combine the Markov chain and stochastic geometry theory. Such analysis of MEC systems can only be addressed by numerical simulations, which is time-consuming and has poor scalability. Fortunately, owing to the recent development of the technology of computation-intensive tasks for edge users to enhance user experience. Note that these two techniques seem to target similar problems but they will be integrated seamlessly in this subsection.

Spatial/Temporal popularity driven caching

• Service Caching

Heterogeneous resources allocation

• Data Caching

CPU-hungry

Memory-hungry

Storage-hungry

Reduce the computation latency

Reduce the computation latency
Mobility Management for MEC

Mobility-aware Offloading Using D2D Communication

Mobility-aware Live Prefetching

Mobility-Aware Server Scheduling

Green MEC: Dynamic Right-Sizing

**Dynamic Right Sizing**

- Energy efficient, but
  - Switching and VM migration cost
  - Degradation of QoS
  - High maintenance costs

*Accurate profile of computation workload at edge cloud should be forecasted.*
Green MEC: Renewable Powered MEC

- Green energy-aware resource allocation & computation offloading
  - Channel & energy & workload information
- Large-scale renewable powered MEC system
  - Spatial diversity of the available renewable energy
Green MEC: Geographical Load Balancing

- **Geographical load balancing for MEC:** leverages the spatial diversities of the workload patterns to make workload routing decision among different edge-clouds.

- **Advantages:**
  - Improve the energy efficiency of the lightly-load edge servers and user experience.
  - Prolong the battery lives of mobiles.
  - Some efficient resource management techniques at edge-cloud are used, such as VM management, dynamic right-sizing, etc.
**Motivation:** Using wireless energy to power MEC

**Objective:** Maximize energy savings given computation deadline constraint

- Offloading or not
- Energy beamforming
- Double near-far problem
- Multiuser scheduling

MEC for 5G
Each edge-cloud server consists of a hosting infrastructure and an application platform.
5G MEC: Challenges and Requirements

- Application Portability
- Network Integration
- Security
- Regulatory and Legal Consideration
- Resilience
- Performance
- Operation

MEC System

Virtualization and Cloud Computing

Optimizing virtualization using advanced memory and storage technology.
5G Applications of MEC

MEC for video stream analysis

MEC for auto-driving

MEC for augmented reality service
5G Specifications for MEC

- Traffic selection
- UPF selection
- Session and service continuity
- Information exchange with AF
- QoS control and charge for routing
How does 5G enable MEC?

- **Support Service Requirement**

  - Latency-sensitive application
  - Mission-critical service
  - Small packet error rate

- **Advanced Mobility Management Strategy (AMMS)**

  5G Core Network → Determine → Mobility Pattern → Decide → AMMS

  Monitor & update
How does 5G enable MEC?

- **Capability of Network Slicing**

![Diagram showing Network Slicing]

- **Business Layer**
  - User and industry business applications
  - Network Function and Application Store

- **Service Layer**
  - Network Applications
  - Unified Management and Orchestration
  - Network Functions
  - Dedicated User Plane

- **Cloud Infrastructure Layer**
  - RRH
  - Network
  - Storage
Is PHY Layer Dead?


Not yet… it is an exciting information age of C’s

• Computing
• Communication
• Control
• Content delivery
• ….
Reference: